RECEIVED
CENTRAL FAX CENTER
JUL 13 2007

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

Claims 1 - 6 (cancel)

Claim 7 (currently amended): A method comprising:

recognizing a face of a subject from first entries in a database, including modeling an image including the face using an embedded hidden Markov model (EHMM), wherein the EHMM is a hierarchical statistical model having a parent layer corresponding to a super state of the EHMM and including a plurality of nodes to represent hidden nodes, each node referring to a plurality of second nodes of a child layer corresponding to a state of the EHMM, the plurality of second nodes each referring to an observation node, and wherein the state of the EHMM is described by a mixture of a plurality of Gaussian density functions having diagonal covariance matrices;

recognizing audio-visual speech of the subject from second entries in the database; and

identifying the subject based on recognizing the face and recognizing the audio-visual speech.

Claim 8 (original): The method of claim 7, further comprising providing the subject access to a restricted area after identifying the subject.

Claim 9 (cancel)

Claim 10 (currently amended): The method of claim [[9]] 7, further comprising obtaining observation vectors from a sampling window of the image.

Claim 11 (original): The method of claim 10, wherein the observation vectors comprise discrete cosine transform coefficients.

Claim 12 (original): The method of claim 7, wherein recognizing the face comprises performing a Viterbi decoding algorithm.

Claim 13 (original): The method of claim 7, wherein recognizing the audio-visual speech further comprises detecting and tracking a mouth region using vector machine classifiers.

Claim 14 (original): The method of claim 7, wherein recognizing the audio-visual speech comprises modeling an image and an audio sample using a coupled hidden Markov model.

Claim 15 (original): The method of claim 7, further comprising combining results of recognizing the face and recognizing the audio-visual speech pattern according to a predetermined weighting to identify the subject.

Claim 16 (currently amended) A system comprising:

at least one capture device to capture audio-visual information from a subject;

a first storage device coupled to the at least one capture device to store code to enable the system to recognize a face of the subject from first entries in a database, model an image including the face using an embedded hidden Markov model, model the image and an audio sample using a coupled hidden Markov model, recognize audio-visual speech of the subject from second entries in the database, and identify the subject based on the face and the audio-visual speech according to a matching score corresponding to $\lambda_f L(O^f|k)+\lambda_{av} L(O^a,O^v|k)$, where O^a, O^v and O^f are audio speech, visual speech and face of the captured audio-visual information, $L(*|k)$ is an observation likelihood for a k^{th} entry in the database, and $\lambda_f \lambda_{av} \geq 0, \lambda_f + \lambda_{av} = 1$ are weighting coefficients for face and audio-visual speech likelihood of recognition; and

a processor coupled to the first storage to execute the code.

Claim 17 (original): The system of claim 16, wherein the database is stored in the first storage device.

Claims 18 – 19 (cancel)

Claim 20 (currently amended): An article comprising a ~~machine~~ computer-readable storage medium containing instructions embodied on the computer-readable medium that [[if]] when executed enable a system to:

recognize a face of a subject from first entries in a database, model an image including the face using an embedded hidden Markov model;

recognize audio-visual speech of the subject from second entries in the database, model the image and an audio sample corresponding to the audio-visual speech using a coupled hidden Markov model; and

identify the subject based on recognizing the face and recognizing the audio-visual speech according to a matching score corresponding to $\lambda_f L(O^f|k)+\lambda_{av}L(O^a,O^v|k)$, where O^a, O^v and O^f are audio speech, visual speech and face of the image and the audio-visual speech, $L(*|k)$ is an observation likelihood for a kth entry in the database, and $\lambda_f\lambda_{av}\geq 0, \lambda_f+\lambda_{av}=1$ are weighting coefficients for face and audio-visual speech likelihood of recognition.

Claim 21 (original): The article of claim 20, further comprising instructions that if executed enable the system to provide the subject access to a restricted area after the subject is identified.

Claims 22 - 23 (cancel)

Claim 24 (new): The method of claim 15, wherein the predetermined weighting corresponds to $\lambda_f L(O^f|k)+\lambda_{av}L(O^a,O^v|k)$, where O^a, O^v and O^f are acoustic speech, visual speech and facial sequence of captured audio-visual information, $L(*|k)$ is an observation likelihood for a kth entry in the database, and $\lambda_f\lambda_{av}\geq 0, \lambda_f+\lambda_{av}=1$ are weighting coefficients for face and audio-visual speech likelihood of recognition.

Claim 25 (new): The article of claim 20, wherein the embedded hidden Markov model (EHMM) is a hierarchical statistical model having a parent layer corresponding to a super state of the EHMM and including a plurality of nodes to represent hidden nodes, each node referring to a plurality of second nodes of a child layer corresponding to a state of the EHMM, the plurality of second nodes each referring to an observation node, and wherein the state of the EHMM is described by a mixture of a plurality of Gaussian density functions having diagonal covariance matrices.